Received: 18 May 2020 | Revised: 30 July 2020 | Accepted: 21 August 2020

DOI: 10.1002/jrs.5984

RESEARCH ARTICLE

Journal of RAMAN SPECTROSCOPY WILEY

Single-cell Raman spectrum extraction from clinic biosamples

Xinxin Han[1,2] | Yihui Wu[2,3] | Mingbo Chi[2] | Sujun Gao[4] | Quan Wang[5]

[1]College of Materials Sciences and Opto-Electronic Technology, University of Chinese Academy of Sciences, Beijing, China

[2]State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, China

[3]Research Center of Materials and Opto-Electronic, University of Chinese Academy of Sciences, Beijing, China

[4]Department of Hematology, The First Bethune Hospital of Jilin University, Changchun, China

[5]Department of Gastroenteric and Anal Surgery, The First Bethune Hospital of Jilin University, Changchun, China

Correspondence
Yihui Wu and Mingbo Chi, State Key Laboratory of Applied Optics, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130031, China.
Email: yihuiwu@ciomp.ac.cn; chimb@sklao.ac.cn

Sujun Gao, Department of Hematology, The First Bethune Hospital of Jilin University, Changchun, China.
Email: sujung1963@sina.com

Funding information
National Natural Science Foundation of China, Grant/Award Number: 6172780105

Abstract

Raman spectra of clinical samples are often affected by both the substrate and background fluorescence. The overlay of the substrate spectrum, biological Raman spectrum, and the fluorescence spectrum significantly affects the identification of biometrics. In this paper, we propose a specific-scale analysis algorithm and a zero-order Savitzky–Golay filtering algorithm combining local minima to separate the substrate and background fluorescent, respectively. The specific-scale analysis algorithm based on wavelet transform can realize the linear separation of the substrate spectrum through multiresolution analysis. The zero-order Savitzky–Golay filter algorithm combining local minima is an empirical algorithm. It estimates the background fluorescence by building a smooth curve that passes through local minima. We tested our algorithms with simulated spectra and with the Raman spectra of clinical biosamples recorded on glass. The analysis results of three gastric cancer cell lines indicated that the classification error of the processed spectrum decreased significantly.

KEYWORDS

background fluorescence, glass substrate, Raman spectrum, specific-scale analysis, zero-order Savitzky–Golay filter

1 | INTRODUCTION

Spontaneous Raman spectroscopy has been widely used in the field of cancer diagnosis.[1–6] Raman spectra based on molecular vibration characteristics can detect changes in molecular components associated with cell carcinogenesis and can achieve accurate and rapid identification of cancerous cells through statistical analysis.[7–11] Compared with surface-enhanced Raman spectroscopy (SERS), spontaneous Raman spectroscopy is highly

repeatable[12]; therefore, it is more suitable for the establishment of a standard library for cancer diagnosis. Moreover, spontaneous Raman spectroscopy has significant clinical application value because it is label-free, nondestructive, and involves lower cost.

However, the Raman spectra of biological samples contain signals representing irrelevant information such as the fluorescence spectrum and substrate spectrum,[13,14] which makes it more challenging to identify the Raman spectra of the trace biomolecules and can have a significant impact on subsequent statistical analysis.[14,15] Glass slides, which are widely used in biology and medicine, produce a strong Raman signal and have a greater impact on the Raman signal of biological samples.[16] It is generally believed that the biological slide raw spectrum S can be divided into three parts, including the real Raman spectrum R of the biological sample, the bioluminescent background signal F, and the glass substrate signal G, regardless of noise.[17–19] However, it is difficult to separate the three unknown spectra owing to the presence of the glass spectrum, which contains both broad-spectrum information similar to fluorescence and narrow-spectrum information similar to a Raman spectrum. Fluorescence occurs concurrently with Stokes Raman scattering upon laser excitation.[20] Mathematical processing is one of the most commonly used fluorescence suppression methods for Raman spectroscopy analysis. It is generally believed that fluorescence changes much more slowly than the Raman signal.[20] The low-frequency spectrum obtained by methods, such as wavelet transformation,[21–23] polynomial fitting,[24,25] morphological analysis,[26–28] and local average,[29–31] is considered to yield the best estimate of the background fluorescence. Mathematical processing is also a feasible method for removing the substrate spectrum. It is a relatively simple method in which the spectra of the glass substrate are collected near the sample and are then subtracted from the raw spectra. However, because of the masking of the sample, the surrounding glass spectrum cannot completely represent the glass substrate below the sample. To separate the three components mixed in the spectrum, it is necessary to introduce reference spectra. Beier and Berger[17] used a polynomial to simulate the fluorescent background and introduced a standardized glass spectrum as the reference spectrum. The coefficient of the reference spectrum and the polynomial were then calculated by solving for the minimum values of the multivariate functions. This method can simultaneously estimate the contributions of the fluorescent and the glass substrate. Kerr et al.[18] proposed an extended multivariate signal correction (EMSC) method that could also achieve separation of the three spectra. The algorithm is implemented via an adapted version of MATLAB's polyfit function. Furthermore, it requires three reference spectra: a standardized glass spectrum, a mean spectrum of the sample without the glass substrate and fluorescence, and a polynomial background. Other methods have also been proposed.[19] However, these methods require a polynomial to simulate the fluorescence background, and it is difficult to find a suitable polynomial that perfectly simulates the background fluorescence in a diverse biological environment. The order and initial value of the polynomial significantly affect the accuracy and speed of the calculations. Therefore, it is necessary to develop an accurate and rapid method to solve this problem.

In this paper, we propose a specific scale analysis (SSA) algorithm and a zero-order Savitzky–Golay filtering algorithm combining local minima (LM-SG) to separate the substrate and fluorescence, respectively. In the SSA algorithm, the raw spectrum and the reference spectrum are decomposed by a fast wavelet transform (WT), and the glass spectrum can be extracted independently by comparing the characteristics of different scales. The LM-SG algorithm is a smoothing method based on the SG filter. The fluorescence can be estimated by constructing a smooth curve across the LM. Both methods were tested using synthetic spectra, simulation spectra, and real Raman spectra, which are widely used in the field of fluorescence removal.[32] The results of principal component analysis (PCA) show that irrelevant information in the spectrum can be ignored after treatment. In addition, the scalability of the SSA algorithm is verified by taking a silica substrate as an example.

2 | METHODOLOGY

2.1 | SSA algorithm

Based on the uniform stability of glass, the glass spectrum G can be represented as the product of a coefficient c and a reference spectrum g. The Raman spectra of clinical samples can then be expressed as

$$S = R + \mathrm{F} + c * g. \tag{1}$$

Figure 1 shows the wavelet-scale characteristics of the three spectra. From the spectral features, we can see that the glass signal contains not only the low-frequency signal close to the fluorescence but also a signal similar to the fluorescent signal and the Raman signal. To extract the real Raman spectrum, it is preferable to start with a minimum number of variables. The decomposition and extraction of a specific spectral scale are effective ways to

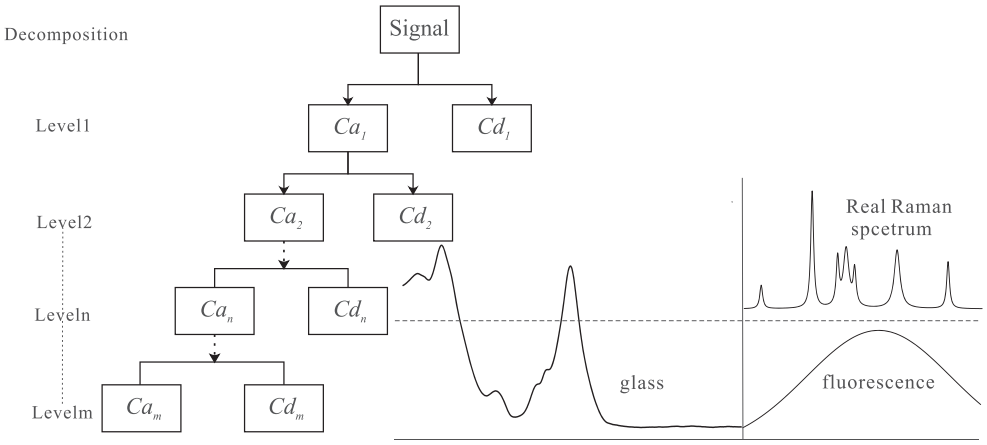

FIGURE 1 The wavelet space corresponding to different spectra (the dashed line represents the scale boundary between the Raman spectrum and the fluorescence; the glass spectrum spans the two scales)

reduce the variables. Multiresolution analysis, also known as multiscale analysis, is an essential basis for the multiscale decomposition of the WT.[33] According to the one-dimensional WT, the raw spectrum S can be represented as

$$S(t) = \sum_k a_{j0}(k)\varphi_{j0,k}(t) + \sum_{j=j_0}^{\infty}\sum_k d_j(k)\psi_{j,k}(t), \qquad (2)$$

where t is the horizontal coordinate of the spectrum (i.e, the wavenumber), k represents the position of the function along the horizontal coordinate, j determines the width of the function, $j0$ is any starting scale, $a_{j0}(k)$ is an approximation or scale coefficient, $d_j(k)$ is the detail or wavelet coefficient, ψ is the basic wavelet or mother wavelet, and $\varphi_{j0,k}(t)$ is the scale function or father wavelet.

Owing to the difference in scale between the real Raman signal R and the background fluorescence F, fluorescence information only appears in large-scale spaces. When decomposed to a certain level n, the influence of the scale coefficient on Raman signal can be neglected. If we remove all scales greater than n in the spectrum, the remaining spectrum contains only the real Raman spectrum R and the partial glass Raman signal G_p. The two signals satisfy the linear superposition relationship:

$$S'(t) = \sum\nolimits_k a_{j0,R}(k)\varphi_{j0,k}(t) + \qquad (3)$$

$$\sum\nolimits_{j=j_0}^{\infty}\sum\nolimits_k d_{j,R}(k)\psi_{j,k}(t) +$$

$$c\left(\sum\nolimits_k a_{j0,pg}(k)\varphi_{j0,k}(t) + \sum\nolimits_{j=j_0}^{\infty}\sum\nolimits_k d_{j,pg}(k)\psi_{j,k}(t)\right).$$

The subscripts R and pg represent the real Raman spectrum and the partial glass Raman spectrum, respectively. Therefore, we only need to decompose both the raw Raman spectrum and the reference spectrum at a specific level of n and extract the components with a scale less than n. Subsequently, we can estimate the coefficient of the glass spectrum by linear fitting of the two extracted spectra. The corresponding decomposition level and the wavelet in the SSA algorithm are discussed in the experimental section.

2.2 | LM-SG algorithm

After the glass substrate is removed, the method of removal of the background fluorescence is a prevalent topic. We tend to set some local minimum points and subtract any smooth polynomial curves that pass through these points to remove the background fluorescence when dealing with a biological Raman spectrum. On this basis, a zero-order SG filtering algorithm combining LM is proposed to estimate the background fluorescence automatically. Iteration of the cubic spline curve is used to find the local minimum, and iteration of the SG filter is used to construct a smooth curve. The SG filter is a typical local polynomial fitting algorithm, and the zero-order SG filter is a good quality smoothing filter.[29,30] The key to estimating the baseline using the SG filter is to set an appropriate frame length. By comparing spline curves with different interpolation intervals, we can quickly identify the local minimum points in the sample. The maximum interval of these local minimum points is the ideal length of the SG filter for baseline estimation. The flow chart of this process is shown in Figure 2.

The calculation steps are as follows:

1. Extract the data points from sample Y at the same interval;
2. The sampling points are fitted using a cubic spline curve, and the size of the samples after fitting should be consistent with the raw spectrum.

3. The fitted result Y_i is compared with the previous result $Y_{i\ \_\ 1}$, and the minimum value of the corresponding position is extracted as the spectrum.
4. When the last two fitting results are completely consistent, the sampling interval i is recorded.
5. Calculate the fluorescence using zero-order SG filtering with a frame length ranging from 2 to i.

In addition, the final curve obtained by spline interpolation fitting can also be regarded as an estimate of the background fluorescence. However, the result is less smooth. While the Raman spectrum with 523-nm excitation light source has a wavelength range of approximately 50 nm in the 400–1,800 cm^{-1} range, the fluorescence spectrum does not fluctuate significantly within this range. Therefore, we chose a relatively smooth zero-order SG filter for fluorescence estimation when extracting a biological Raman spectrum.

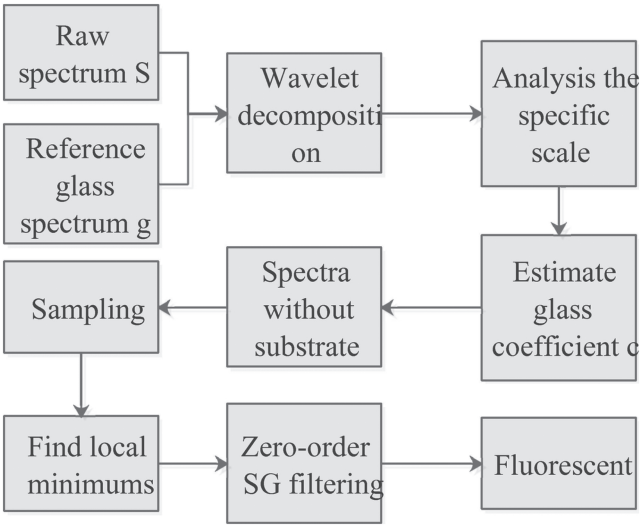

FIGURE 2 The flow chart of SSA and LM-SG algorithm

3 | SIMULATED AND EXPERIMENTAL SPECTRA

The two algorithms were tested using synthetic spectra, simulation spectra, and real Raman spectra. It is assumed that any noise in the spectra has been adequately processed.

3.1 | Raman spectroscopy system

The Raman spectrum was collected using a Raman spectroscopy system built in our lab. The system uses a 523-nm excitation light source (200 mw). The laser is focused through a 50X microscope objective (Olympus) into a 2-μm spot. An Andor CCD camera is used to detect the Raman signal after dispersion. The Raman spectra covering the 400–1,800 cm^{-1} spectral range were recorded, and the resolution of the spectrometer was approximately 2 cm^{-1}. In our study, a neutral density filter (THORLABS NE06B) was used to attenuate the laser intensity, and the exposure time was 20 s. The schematic of the experimental system is shown in Figure 3. All cell lines used in our experiments were provided by the First Bethune Hospital of Jilin University.

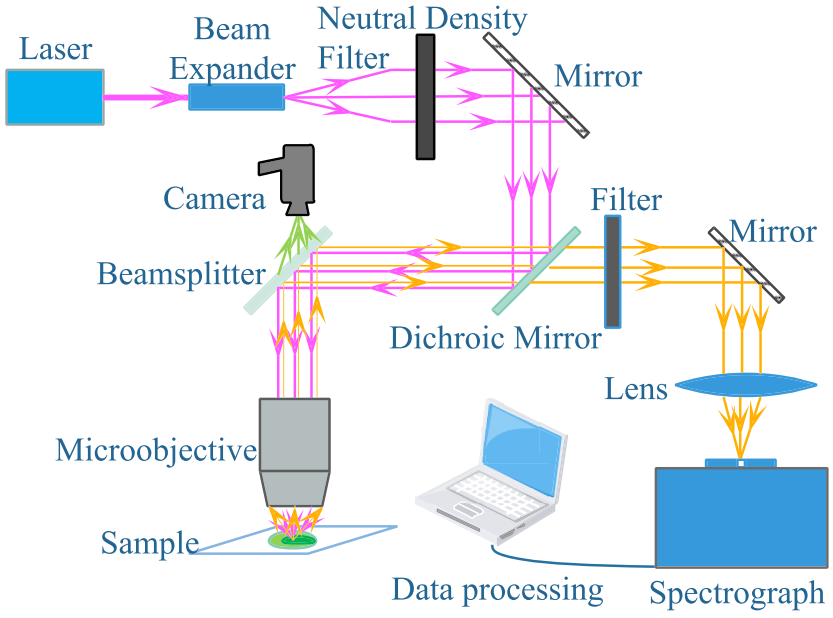

FIGURE 3 The scheme of the experimental system [Colour figure can be viewed at wileyonlinelibrary.com]

3.2 | Synthetic spectra

In this study, seven Lorenz peaks were used to simulate the Raman spectrum,[32] and the expressions of the Lorenz curve are as follows:

$$y = \frac{a}{b^2 + (x-c)^2},$$

where $\frac{a}{b^2}$ is the peak height, b is the half-height width, and c is the position of the peak.

The parameters of the synthetic Raman spectrum are listed in Table 1.

Flat, linear, Gaussian, and sigmoidal curves were used to simulate the background fluorescence.

The flat background can be represented by any constant, which was set to 0 in this study.

The expression for the linear background is

$$B_l = \frac{x}{50}.$$

The expression for the Gaussian background is

$$B_g = 10e^{-\left(\frac{x-1200}{600}\right)^2}.$$

The expression for the sigmoidal background is

$$B_s = \frac{20}{1 + e^{-0.01(x-1100)}},$$

where B is the intensity of the background fluorescence and x is the Raman shift.

The mean spectrum of 10 slides was used as the glass substrate spectrum. The synthetic spectra can be used to evaluate the adaptability of the two algorithms to different fluorescent backgrounds and the influence of the two parameters in the SSA algorithm. The results are shown in Figure 4.

TABLE 1 The parameters of Lorenz curve used for constructing the synthetic Raman spectrum

a	400	2,000	800	4,000	600	4,000	800
b	10	10	10	20	10	20	10
c	500	800	950	1,000	1,050	1,300	1,600

3.3 | Simulation spectra

To demonstrate the effectiveness of the SSA and LM-SG algorithms in practical applications, we tested both

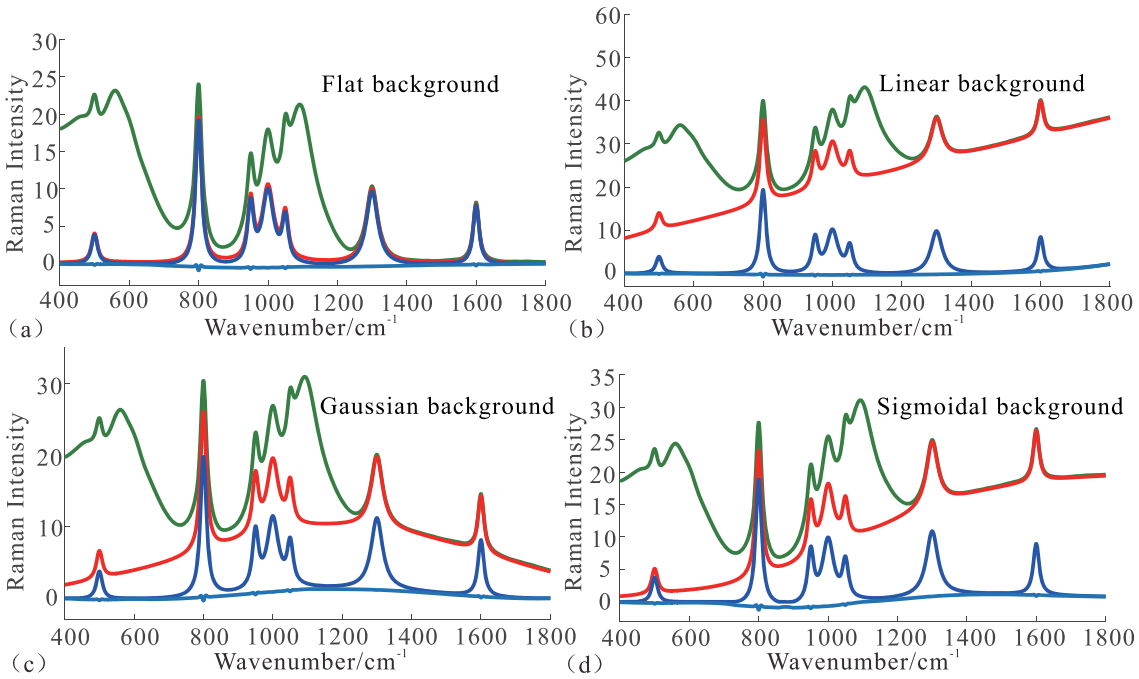

FIGURE 4 Processing results of simulated spectra with different fluorescence backgrounds: (a) flat background, (b) linear background, (c) Gaussian background, (d) sigmoidal background; four colors, respectively, represented the raw spectrum (green), the removed glass substrate spectrum (red), the recovered spectrum (blue), and the residual between the recovered spectrum and the synthetic Raman spectrum (cyan) [Colour figure can be viewed at wileyonlinelibrary.com]

algorithms using the simulation spectrum containing a fetal bovine serum's Raman spectrum and a glass substrate's spectrum. The fetal bovine serum was placed on the surface of a calcium fluoride substrate. Calcium fluoride, an alternative substrate material in Raman research, exhibits a feature spectrum before 400 cm^{-1} and a weak Raman signal from 500 to 700 cm^{-1}. The calcium fluoride substrate spectrum cannot be ignored when the sample is thin. In our experiments, this phenomenon was not considered because of the high concentration and thickness of the serum. The spectra of the serum and glass substrate were collected, as shown in Figure 6a. A third-order SG filter was used to remove the noise in the two spectra, and the frame length was 15. The glass spectrum was normalized and multiplying by 500 and was added to the Raman spectrum of the serum. The spectrum was processed by our method and the Beier method, respectively.

3.4 | Experimental spectra

The goal of this study was to extract the Raman spectra of the biological components of clinical samples such as tissue slices and cell smears; therefore, testing with biological samples was considered essential. In this study, 59 spectra of three different gastric cancer cell lines (BGC823, MGC803, and SGC7901) and 60 spectra of leukemic cell lines (HL60 and Jurkat) recorded on glass slide were tested. The spectra of each slide were probed at three different random points, and the average was used as the reference spectrum. All spectra of samples were processed using the combined SSA and LM-SG algorithms and the Beier method. Subsequently, the raw spectra and processed spectra were, respectively, analyzed by PCA and linear discriminant analysis (LDA), which are widely used for the cluster analysis of Raman spectra.[16,34]

4 | RESULTS AND DISCUSSION

4.1 | Synthetic spectra

Figure 4 demonstrates the adaptability of the two algorithms to different fluorescent backgrounds. The two algorithms can separate the three spectra well under different fluorescent backgrounds. The absolute value of residuals between the recovered spectrum and the synthetic Raman spectrum is small, and most values are less than 0.1, which is much lower than the intensity of the Raman peaks. The SSA algorithm is not affected by the fluorescence-removed method applied. SSA can individually remove the glass substrate and in combination with any de-fluorescence method, which provides more possibilities for subsequent improvement.

4.2 | Wavelet and decomposition level

The key to applying the SSA algorithm is to find the appropriate wavelet and decomposition level. To explore the influence of these two parameters, we selected different wavelets and different decomposition levels to process the Raman spectra with four different backgrounds. The root mean square error (RMSE) value between the processed spectrum and the ideal spectrum was then calculated. The results are shown in Figure 5, and the RMSE is calculated as

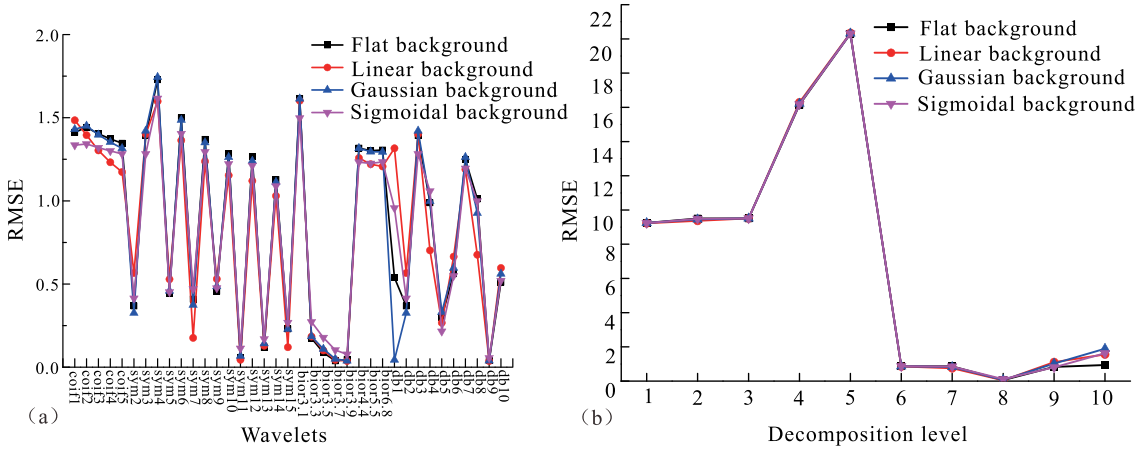

FIGURE 5 (a) The RMSE of the four backgrounds under different wavelet when the decomposition level is 8; (b) the RMSE of the four backgrounds under different decomposition level when the wavelet is 'sym11' [Colour figure can be viewed at wileyonlinelibrary.com]

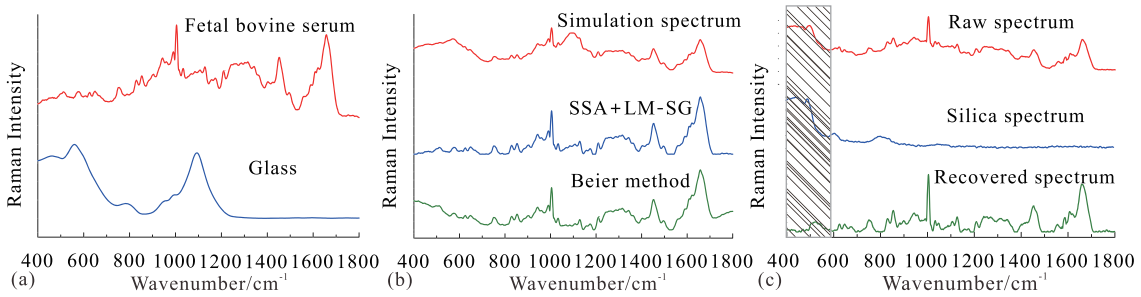

FIGURE 6 (a) Fetal bovine serum Raman spectrum on calcium fluoride and a glass reference spectrum; (b) the simulation spectrum and the processed results by SSA and LM-SG algorithm and Beier method, respectively; (c) Raman spectrum of fetal bovine serum on silica and the recovered spectrum processed by SSA and LM-SG algorithm; the shadow portion represents the feature segment of the substrates [Colour figure can be viewed at wileyonlinelibrary.com]

$$RMSE = \sqrt{\frac{\sum_n (y_c - y_a)^2}{n}},$$

where y_c and y_a represent the recovered spectrum and the real spectrum, respectively, and n is the length of the spectrum.

The Daubechies family wavelets are commonly used in background elimination methods.[35–37] The Symlet family wavelet is developed based on the Daubechies family wavelet; it has better symmetry and can reduce the phase distortion in the analysis and reconstruction of signals. Figure 5 shows that the Raman spectra of four types of fluorescent backgrounds can be better recovered using the 'sym11' wavelet when the decomposition level is 8. When the decomposition level is 1–3, the RMSE values are almost unchanged. This indicates that all Raman signals are completely removed at this time; the remaining spectra are only noise components, and the results are meaningless. With an increase in the decomposition level, part of the biological Raman signal is retained, and the glass spectrum is completely removed. However, in this case, the reference spectrum is not valid; therefore, the error is large. When the level is increased to 8, the fluorescent background is completely removed, and the remaining spectrum contains only the real Raman spectrum and the glass spectrum corresponding to the reference spectrum. Therefore, the glass coefficient can be easily obtained by linear fitting. The analysis results for the different wavelets and decomposition levels also indicate that the SSA algorithm is not affected by fluorescent background.

4.3 | Simulation spectra

In Figure 6, the glass coefficient estimated by the SSA algorithm is 499.8055, which is close to the actual value of 500. The glass coefficient estimated by the Beier method is 610.8831, which is significantly larger than the actual value. The error is caused by the order and initial value of the polynomial. The result obtained by the Beier method has obvious characteristics of second-order polynomials, while the real spectrum was completely recovered in the specific region of its glass spectrum when the SSA and LM-SG algorithms were applied. This indicates that the SSA algorithm can fully meet the analysis requirements for the actual detection and has strong adaptability to the complex background fluorescence in biological samples. In addition, the SSA and LM-SG algorithms can be used not only for the Raman spectra of glass materials but also for other materials with similar characteristics. The silica Raman spectrum, a common interference factor in fiber Raman spectroscopy, dramatically affects the application of fiber Raman spectra in vivo detection.[19,38] To explore the treatment capacity of our method regarding biological samples on a silica substrate, we placed 50 μl of fetal bovine serum on the surface of a silica substrate and collected the Raman spectrum and substrate spectrum after drying naturally. Subsequently, the SSA and LM-SG algorithms were used to remove the substrate and background fluorescence. The results are shown in Figure 6c. After applying SSA and LM-SG algorithms, the silica substrate spectrum (shaded region) was almost completely removed. Therefore, we can conclude that the SSA algorithm also has practical value in the background correction of fiber Raman spectroscopy. Provided that the substrate material is uniform and stable, the SSA algorithm can remove the substrate spectra contained in the Raman spectra of biological samples.

4.4 | Clustering analysis

We used the common PCA method to reduce the dimensions of the spectra of the three cell lines. The first three PC scores were tested by analysis of variance

(ANOVA) (a type of variance analysis widely used in statistical analysis), and the statistical differences of the first three PCs were significant ($P < .01$). Subsequently, LDA was used to classify further the first three PCs (Figure 7), and 10-fold cross-validation was used to estimate the classification error. The spectrum processed by the SSA and LM-SG algorithms exhibits a visible clustering distribution having clear boundaries and can be effectively distinguished. The influence of glass signals at 1,000–1,200 cm^{-1} can be ignored, as shown in Figure 7e. In contrast, it is difficult to separate the three cell lines when the spectra are unprocessed or are processed by the Beier method. The PC coefficients of the first three PCs indicate that glass signals in the untreated spectra and the spectra processed by the Beier method have a significant impact on the results of the PCA. Because the second-order polynomial is used to simulate the fluorescent background in the Beier method, obvious characteristics of the glass spectrum and the second-order polynomial are present in the PC coefficients. This dramatically increases the analysis error. The classification error of the spectra processed by our algorithm is 13.56%, which is less than the 37.29% and 42.37% errors achieved by the unprocessed spectra and the Beier treatment, respectively. In the classification of multiple samples, the linear discriminant method may not accurately describe the boundary, resulting in a large classification error. Therefore, a forward neural network was used for further verification. The results show that the classification error of the spectrum processed by our algorithm is only 1.67%, whereas the classification errors using the Beier method and without processing are 25% and 22.33%, respectively. The results of the two different classification methods indicate that the glass Raman signal and background fluorescence in the biological sample spectrum can be effectively removed by using the proposed SSA and LM-SG algorithms and can effectively improve the accuracy of Raman spectral classification.

We further tested the stability of the two algorithms when the Raman signal was weak. In this experiment, a set of leukemic cell lines where glass signals and fluorescent backgrounds were much stronger than the Raman signals were tested. The results are shown in Figure 8. In the untreated spectrum, the glass and fluorescence signals are the main components of the spectral information and have a dominant influence on PCA. Although classification is apparently possible in this case, the uncertain effects of background spectra are not desirable. The spectral PCs processed using SSA and LM-SG contain almost no background interference, and we can clearly distinguish the Raman spectrum peaks near 1,095 and 1,128 cm^{-1}, which represent distinct types of leukemia.5, 39 This is the desired result.

4.5 | Method comparison

To intuitively illustrate the advantages of the SSA and LM-SG algorithms, we compared the RMSE values of the synthetic spectra processed by several conventional

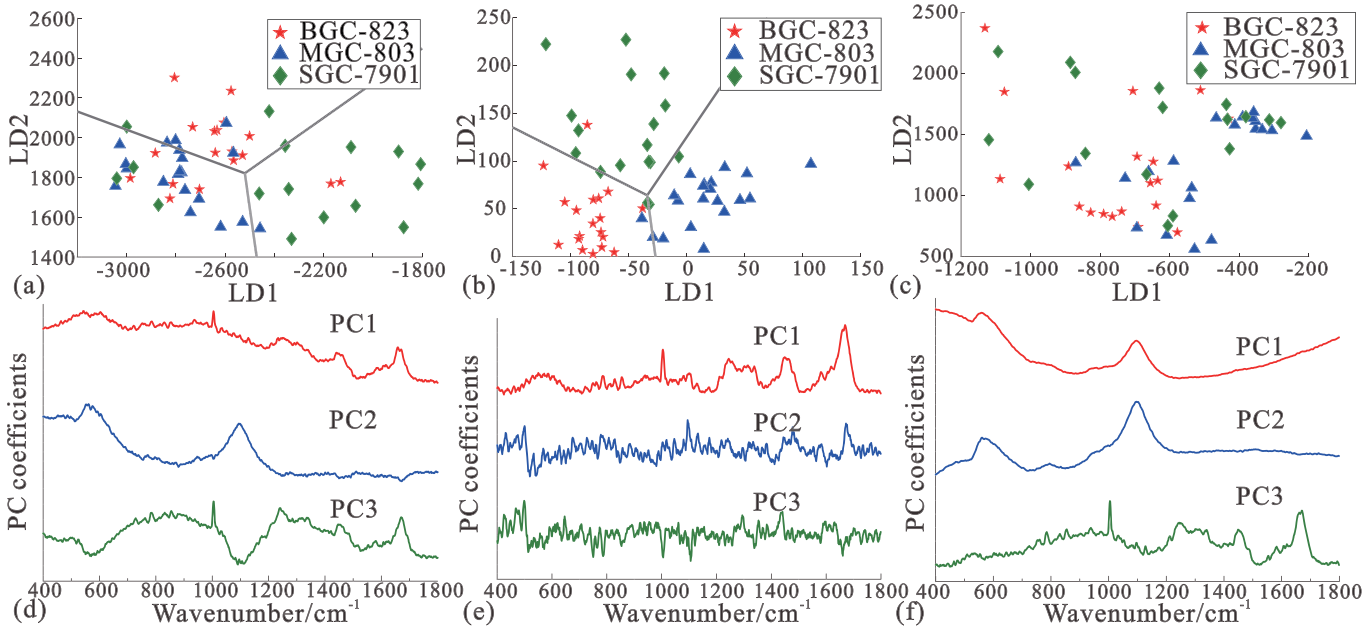

FIGURE 7 The PCA-LDA classification results (the gray line is the classification line) and their corresponding principal component coefficients of different processing methods. (a,d) Remove noise only; (b,e) processed by SSA + LM-SG algorithm; (c,f) processed by Beier method [Colour figure can be viewed at wileyonlinelibrary.com]

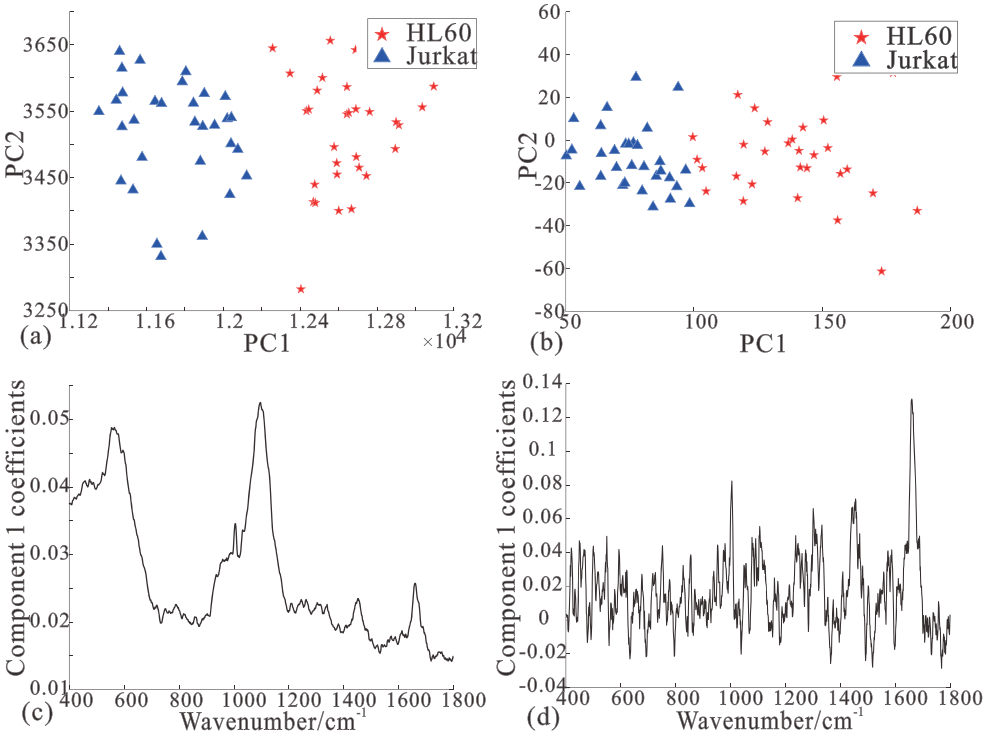

FIGURE 8 Principal component distribution and principal component 1 coefficient of two leukemia cell lines treated by different methods. (a,c) Remove noise only; (b,d) processed by SSA + LM-SG algorithm [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 The RMSE values of the synthetic spectra processed by several common methods

	Flat	Linear	Gaussian	Sigmoidal
SSA + LM-SG	0.3434	0.5596	0.6972	0.7418
Beier	4.5605	4.7237	4.2229	4.9339
airPLS	3.0742	1.9145	2.1305	1.9149
SSA + airPLS	0.5648	1.6022	0.788	1.0089
SSA + WT	2.3141	2.3242	2.3174	2.315

Abbreviations: airPLS, adaptive iteratively reweighted penalized least squares; LM-SG, zero-order Savitzky–Golay filtering algorithm combining local minima; RMSE, root mean square error; SSA, specific scale analysis; WT, wavelet transform.

methods, as shown in Table 2. In this section, we introduce the adaptive iteratively reweighted penalized least squares (airPLS)[31] method and the basic WT to remove fluorescent backgrounds. The polynomial order in the Beier method and airPLS is set to 2.[17,18,31] The parameters in the WT are the same as in SSA. The RMSE values indicate that the glass and fluorescent components in the spectrum cannot be completely removed using only the de-fluorescence method. After removing the glass substrate spectrum using the SSA algorithm, the fluorescent processing algorithm can then significantly remove the background fluorescence, and the LM-SG algorithm has excellent advantages over traditional polynomial fitting and wavelet transformation methods in removing fluorescent backgrounds. Thus, SSA and LM-SG algorithms are demonstrated to be practical and significant for glass substrate removal and background fluorescence correction. We further evaluated the running time of the SSA and Beier methods. A total of 59 samples were used in the experiment, and the final average running time was obtained after five runs on each sample. The average run time of the SSA algorithm was 6.19 s, which is considerably shorter than the 15.09 s required by the Beier method.

5 | CONCLUSION

The SSA and LM-SG algorithms can effectively remove the substrate spectrum and fluorescence spectrum in clinical biological samples. The algorithms are not affected by the initial value and are highly adaptable to different types of fluorescent backgrounds. Compared with the raw spectrum, the spectrum after removing the substrate and the fluorescent background spectra is more conducive to the subsequent analysis of biological samples when using Raman spectroscopy. The precise processing of Raman spectra of clinical samples has a major positive influence on the diagnosis and classification of diseases. These two methods are simple to operate, which makes them an effective pretreatment method for clinical Raman diagnosis. Accurate extraction of biological samples of Raman spectra allows us to

perform rapid Raman testing on larger numbers of clinical samples and provides broader application prospects for Raman spectroscopy.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant number, 6172780105) and Bethune Medical Engineering and Instrumentation Center Foundation Project (Grant number, BQEGCZX2019001).

ORCID

Xinxin Han https://orcid.org/0000-0001-5223-8254
Mingbo Chi https://orcid.org/0000-0001-8683-4257

REFERENCES

[1] M. Li, S. R. Banerjee, C. Zheng, M. G. Pomper, I. Barman, *Chem. Sci.* **2016**, *7*, 6779.
[2] R. E. Kast, S. C. Tucker, K. Killian, M. Trexler, K. V. Honn, G. W. Auner, *Cancer Metastasis Rev.* **2014**, *33*, 673.
[3] P. Crow, B. Barrass, C. Kendall, M. Hart-Prieto, M. Wright, R. Persad, N. Stone, *Br. J. Cancer* **2005**, *92*, 2166.
[4] E. M. Barroso, R. W. H. Smits, T. C. Bakker Schut, I. Ten Hove, J. A. Hardillo, E. B. Wolvius, R. J. Baatenburg De Jong, S. Koljenović, G. J. Puppels, *Anal. Chem.* **2015**, *87*, 2419.
[5] S. Managò, G. Zito, A. C. De Luca, *Optics & Laser Technology.* **2018**, *108*, 7.
[6] S. Mattana, M. Mattarelli, L. Urbanelli, K. Sagini, C. Emiliani, M. D. Serra, D. Fioretto, S. Caponi, *Light: Science & Applications.* **2018**, *7*, 17139.
[7] E. Widjaja, W. Zheng, Z. Huang, *Int. J. Oncol.* **2008**, *32*, 653.
[8] P. Chen, R. Shimada, S. Yabumoto, H. Okajima, M. Ando, C. Chang, L. Lee, Y. Wong, A. Chiou, H. Hamaguchi, *Sci. Rep.-UK* **2016**, *6*.
[9] I. I. Patel, F. L. Martin, *The Analyst.* **2010**, *135*, 3060.
[10] S. Li, Y. Zhang, J. Xu, L. Li, Q. Zeng, L. Lin, Z. Guo, Z. Liu, H. Xiong, S. Liu, *Appl. Phys. Lett.* **2014**, *105*, 91104.
[11] D. Bury, C. L. M. Morais, M. Paraskevaidi, K. M. Ashton, T. P. Dawson, F. L. Martin, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2019**, *206*, 89.
[12] C. Zong, M. Xu, L. Xu, T. Wei, X. Ma, X. Zheng, R. Hu, B. Ren, *Chem. Rev.* **2018**, *118*, 4946.
[13] A. N. Kuzmin, A. Pliss, A. V. Kachynski, *J. Raman Spectrosc.* **2013**, *44*, 198.
[14] H. J. Byrne, P. Knief, M. E. Keating, F. Bonnier, *Chem. Soc. Rev.* **2016**, *45*, 1865.
[15] E. M. Barroso, T. C. Bakker Schut, P. J. Caspers, I. P. Santos, E. B. Wolvius, S. Koljenović, G. J. Puppels, *J. Raman Spectrosc.* **2018**, *49*, 699.
[16] D. W. Shipp, F. Sinjab, I. Notingher, *Adv. Opt. Photonics.* **2017**, *9*, 315.
[17] B. D. Beier, A. J. Berger, *The Analyst.* **2009**, *134*, 1198.
[18] L. T. Kerr, B. M. Hennelly, *Chemometr. Intell. Lab.* **2016**, *158*, 61.
[19] J. Huang, T. Shi, B. Gong, X. Li, G. Liao, Z. Tang, *Appl. Spectrosc.* **2018**, *72*, 1632.
[20] D. Wei, S. Chen, Q. Liu, *Appl. Spectrosc. Rev.* **2015**, *50*, 387.
[21] A. Górski, F. Ciepiela, M. Jakubowska, W. W. Kubiak, *Electroanalysis* **2011**, *23*, 2658.
[22] C. Camerlingo, F. Zenone, G. M. Gaeta, R. Riccio, M. Lepore, *Meas. Sci. Technol.* **2006**, *17*, 298.
[23] M. Liu, Z. Dong, G. Xin, Y. Sun, R. Qu, *Chemometr. Intell. Lab.* **2018**, *182*, 1.
[24] C. A. Lieber, A. Mahadevan-Jansen, *Appl. Spectrosc.* **2003**, *57*, 1363.
[25] H. Hu, J. Bai, G. Xia, W. Zhang, Y. Ma, *Photonic Sensors.* **2018**, *8*, 332.
[26] M. Koch, C. Suhr, B. Roth, M. Meinhardt Wollweber, *J. Raman Spectrosc.* **2017**, *48*, 336.
[27] J. J. González-Vidal, R. Pérez-Pueyo, M. J. Soneira, *J. Raman Spectrosc.* **2017**, *48*, 878.
[28] Y. Chen, L. Dai, *Appl. Spectrosc.* **2018**, *72*, 731.
[29] H. G. Schulze, R. B. Foist, K. Okuda, A. Ivanov, R. F. B. Turner, *Appl. Spectrosc.* **2012**, *66*, 757.
[30] H. G. Schulze, R. B. Foist, K. Okuda, A. Ivanov, R. F. B. Turner, *Appl. Spectrosc.* **2011**, *65*, 75.
[31] Z. Zhang, S. Chen, Y. Liang, *The Analyst.* **2010**, *135*, 1138.
[32] H. Krishna, S. K. Majumder, P. K. Gupta, *J. Raman Spectrosc.* **2012**, *43*, 1884.
[33] S. G. Mallat, *IEEE T. Pattern Anal.* **1989**, *11*, 674.
[34] T. Ichimura, L. Chiu, K. Fujita, H. Machiyama, T. Yamaguchi, T. M. Watanabe, H. Fujita, *Sci. Rep.-UK* **2016**, *6*.
[35] Y. Hu, T. Jiang, A. Shen, W. Li, X. Wang, J. Hu, *Chemometr. Intell. Lab.* **2007**, *85*, 94.
[36] C. M. Galloway, E. C. Le Ru, P. G. Etchegoin, *Appl. Spectrosc.* **2009**, *63*, 1370.
[37] J. Li, L. Choo-Smith, Z. Tang, M. G. Sowa, *J. Raman Spectrosc.* **2011**, *42*, 580.
[38] I. Latka, S. Dochow, C. Krafft, B. Dietzek, J. Popp, *Laser Photonics Rev.* **2013**, *7*, 698.
[39] J. W. Chan, D. S. Taylor, S. M. Lane, T. Zwerdling, J. Tuscano, T. Huser, *Anal. Chem.* **2008**, *80*, 2180.

How to cite this article: Han X, Wu Y, Chi M, Gao S, Wang Q. Single-cell Raman spectrum extraction from clinic biosamples. *J Raman Spectrosc*. 2020;51:2255–2264. https://doi.org/10.1002/jrs.5984